

# ソーシャルメディアにおける情報の伝播がもたらす商品の売れ行きに対する影響について

～ Twitter のつぶやき数の推移と iTunes AppStore のモバイルアプリの売れ行きの相関性 ～

奥田 輔<sup>†</sup> 安田 孝美<sup>†</sup> 水野 政司<sup>††</sup>

<sup>†</sup> 名古屋大学大学院 情報科学研究科 社会システム情報学専攻

<sup>††</sup> クエリーアイ株式会社

**あらまし** 近年、ソーシャルメディアの普及に伴い、ユーザが情報の発信、受信の双方を手軽に、かつリアルタイムな情報の共有ができるようになった。本研究ではそのようなソーシャルメディアのひとつとして、マイクロブログサービス「Twitter」を取り上げ、その中で特定の商品に対するつぶやき数の推移とその商品の売れ行きとの間にいかなる相関が見て取れるかを検証した。今回、つぶやきの集計期間は2011年1月1日から2011年3月19日までの78日間、集計対象とした商品は iTunes App Store で有料にて販売されている iPhone・iPod touch・iPad 対応のモバイルアプリ、1235 種類である。統計の結果、Twitter のつぶやき数ランキング、iTunes App Store のランキングの双方が24時間以内に、1時間あたり10位以上のランク上昇が見られる割合は、全体の49.79%となった。また、つぶやき数ランキングが上昇してからそれに追隨して iTunes App Store のランキングが上昇する割合は37.17%、その時間遅れは平均値で4.9時間後、中央値で6時間後にする傾向が見られた。

**キーワード** ツイッター, ソーシャルメディア, モバイルアプリ, 統計, 販売戦略

## Research into the impact on sales caused by the spread of information on SNS

～ Correlation between number of tweets on Twitter and sales of mobile application on iTunes AppStore ～

Tasuku OKUDA<sup>†</sup>, Takami YASUDA<sup>†</sup>, and Masashi MIZUNO<sup>††</sup>

<sup>†</sup> Department of Systems and Social Informatics, Graduate School of Information Science, Nagoya University

<sup>††</sup> QueryEye, Inc.

**Abstract** With the recent spread of Social Media, people can share information easily and even in real-time. In this study, we investigated the correlation between the number of tweets in the twitter about a product such as an application on iTunes App Store and its sales. As a result of the study, the followings are found out. The percentage of rapid ranking rise in both the tweets and iTunes App Store was 49.79%. The case in which the boom in iTunes App Store followed rapid increase in the ranking of the tweets was 37.17%, and its average time lag was 4.9hours, the median value was 6hours.

**Key words** Twitter, SNS, Mobile Application, Statistics, Sales Strategy

### 1. 研究の背景

昨今のインターネット上における情報交換ツールとして、ソーシャルメディアの重要性は非常に大きい。その利用形態は特定の話題について議論する電子掲示板や情報発信ツールとしてのブログ、ソーシャル・ネットワーキング・サービス（以下、SNS）など様々である。これらの普及の背景には、その利用の手軽さとともにメディアでの認知や有名人の利用、モバイルア

プリケーションの整備などが挙げられる。またこれらツールを利用することで誰もが情報の発信、受信の双方を手軽に行えるようになる。最近では SNS と呼ばれるサービスが利用者数を急激に伸ばしており [1]、従来のブログや掲示板といった特定の話題に沿って情報のやりとりを行うのではなく、人と人との繋がりによって情報を共有する機会が増えてきている。

このようなソーシャルメディアにて最近取り入れられている一つの特徴として、自身から情報を発信するだけでなく、情報

の共有が比較的容易に行える、またはそのような機能を備えているという点がある。例を挙げると、マイクロブログサービス「Twitter」でのリツイート、SNS サービス「Facebook」でのいいね！ボタンなどである。これらの機能はユーザが価値があると判断した情報を自身に近いユーザと共有することができる機能である。この機能は、単純に他者にその情報を知らせるだけでなく、その情報を他者と共有することでその情報に触れることができるユーザを増やし、情報の価値を高める行為と捉えることができる。

本研究では、そのようなソーシャルメディアが社会に与える影響を調べるために、マイクロブログサービス「Twitter」に着目し、その中で後述するモバイル端末向けアプリケーションに関するつぶやき数を抽出したランキングと、Apple 社のモバイル端末 iPhone・iPod touch・タブレット端末 iPad 向けアプリケーションのダウンロードサービスである「iTunes App Store」のダウンロード数ランキングを統計、比較しその間にいかなる相関が存在するかを調査した。

## 2. 関連研究

ソーシャルメディアとしての Twitter に関する初期の論文として、Java らの論文 [2] がある。この研究では、Twitter の利用者動向や世界的な分布、フォロワー数・フォロー数との関係などを調査している。また、ユーザの行動を分析し、ユーザがいかにして Twitter を利用しているかを述べている。それによると、ユーザは他のユーザとのコミュニケーションや、自身の興味ある分野の情報の収集、また日記のような他愛もないつぶやきをメインとして投稿していることが挙げられている。そして、Hypertext Induced Topic Selection アルゴリズムを用いて、ユーザの重み付けからコミュニティがいかに形成されているかを分析している。

ソーシャルメディアが経済に与えるを調べたものとして、Bollen らの調査 [3] が挙げられる。この調査では、Twitter のつぶやき本文を OpinionFinder を用いて Positive, Negative の 2 種類、更には Google Profile of Mood States を用いて Calm, Alert, Sure, Vital, Kind, Happy の 6 種類によって解析し、NY ダウ平均株価の上下予想を行った。また、その結果を Bivariate Granger causality Analysis, Self-Organizing Fuzzy Neural Network モデルにより適応させた。その結果、Self-Organizing Fuzzy Neural Network モデルによる Calm の感情抽出値を用いることで、NY ダウ平均株価を 87.6%の精度で予想することに成功した。

ユーザがソーシャルメディアに投稿した情報を元に情報推薦を行う研究としては、白木らの研究 [4] では、Twitter の発言中から「なう」という単語を含む発言を自然言語処理を用いて分析し、発言者の状況を検出、推定することで、発言者の状況に即したモバイルアプリケーションを推薦するシステムを構築している。

## 3. 定義

本章では、今回統計処理をするにあたって用いた種々の定義

を述べる。

### 3.1 つぶやき

今回集計したつぶやきは、各言語のパブリックタイムライン [5] に表示されるツイートを対象に以下の条件を満たすものを集計している。

- つぶやき本文中に、あるアプリケーションの固有 ID を含む iTunes App Store への URL が存在する。
- つぶやき本文中に、あるアプリケーションの固有 ID を含む iTunes App Store へのリンクを含むブログ記事の URL が存在する。

これに該当するつぶやきの投稿された時間、本文、投稿者の Twitter アカウント名、つぶやかれたアプリケーションの固有 ID をデータベースに集計した。

本研究にて集計した期間は 2011 年 1 月 1 日～2011 年 3 月 19 日の 78 日分である。また、調査対象とした iTunes App Store のカテゴリはモバイルアプリ・有料総合である。ここでモバイルアプリとは、対象とする機器が Apple 社のモバイル端末である iPhone・iPod touch、タブレット端末の iPad に対応しているアプリケーションのことを指す。

### 3.2 iTunes App Store

iTunes Search API [6] を用いて、毎日各カテゴリごとに 1 位～301 位までを xml 形式にて問い合わせ、集計した。

本研究にて集計した期間は 3.1 と同じく 2011 年 1 月 1 日～2011 年 3 月 19 日の 78 日分、調査対象とした iTunes App Store のカテゴリはモバイルアプリ・有料総合である。

### 3.3 急上昇

本研究における各ランキングにおける「急上昇」は、最小二乗法を用いて 6 時間毎に直線近似しその傾きが  $-10$  以下となるもの、即ちおしなべて 1 時間あたり 10 位以上の順位上昇がみられるものを急上昇と定義した。

なお、ランキングの性質上、ランクが上昇するとランクの数値自体は減少する。ゆえに、直線近似を行った場合、その傾きが負数である場合がランキングの上昇であり、その絶対値が大きければ大きいほど急上昇と捉えることができる。

### 3.4 相関と因果

まずは iTunes App Store ランキングと実ダウンロード数の間の相関関係について説明する。これは、iTunes App Store のランキングアルゴリズムが後述する式 (1) を用いられていると仮定すれば、近い時間における実ダウンロード数が増えればランキングスコアも増加し、それに伴いランキングも上昇する、と考えられる。これにより iTunes App Store ランキングと実ダウンロード数には、強い相関があると思われる。

次に、今回対象とするつぶやき数ランキングとアプリケーションの参照数についての相関関係について考える。あるアプリケーションの固有 ID を含む URL が書かれたつぶやきが増加することで、その URL が参照される確率が高まる。

本研究では、この「ダウンロード数が増えれば、iTunes App Store のランキングが増加する」という因果と、「アプリケーションへの URL を本文中に含む Twitter のつぶやき数が増加することで、そのアプリケーションへの参照機会が増える」という

因果の両者について、その間にある相関関係を明らかにすることを目的としている。

なお、本研究では「Twitter のつぶやき数増加が iTunes App Store のアプリケーションの売上を向上させる」という因果関係を証明することを目的としていない。

## 4. 計算方法

### 4.1 つぶやき集計

3.1 で述べたとおり、本研究ではつぶやき本文中にあるアプリケーションの固有 ID を含む iTunes App Store への直接のリンク、もしくはブログ記事を介してのリンクがあるつぶやきをカテゴリ、アプリケーションの固有 ID 別に 1 時間ごとに集計した。

### 4.2 つぶやきのランキング化

次に、アプリケーションごとのランキングを作成する。ただし 4.1 で集計した数値は 1 時間ごとの数値であり、非常に落差が激しく、ランキング化に適さない。そこで以下の数式を用いて時間によって重み付けされたスコアを算出した。

$$S_i = \sum_{j=0}^{23} n_{i-j} * 8 + \sum_{j=24}^{47} n_{i-j} * 5 + \sum_{j=48}^{71} n_{i-j} * 5 + \sum_{j=72}^{95} n_{i-j} * 2 \quad (1)$$

ここで  $S_i$  は時間  $i$  におけるスコアを、 $n_i$  は時間  $i$  におけるつぶやき数を表す。式 (1) は iTunes App Store にて売上をランキングのスコアにする際に用いられていると推測される数式である [7]。この数式は、Apple 社の公式な発表によるものではないものの、実際のモバイルアプリケーション [8] の販売実績に基づくもの [9] であり、非常に信頼できる数式と考えられる。このスコアをつぶやき数にも適用し、1 時間ごとにランキングを算出した。

式 (1) を用いて、Twitter のつぶやき数、及び iTunes App Store の順位付けをしたグラフの例を図 1 に示す。ここでは Apple 社の iPhone, iPod touch, iPad 用アプリケーション、“iMovie”を例としてあげる。実線が Twitter のつぶやき数ランキングの推移、破線が iTunes App Store の有料総合カテゴリにおけるアプリケーションの推移である。

### 4.3 急上昇検出

ランキングのグラフを最小二乗法を用いて直線近似した。

#### 4.3.1 最小二乗法

ある測定で得られた数値の組の集合  $(x, y) = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  があるとする。これらの  $(x, y)$  の分布が、ある関数  $y = f(x)$  に従うと仮定した場合、 $x_i$  とそれに対応する想定される理論値の組は  $(x, y) = (x_1, f(x_1)), (x_2, f(x_2)), \dots, (x_n, f(x_n))$  となる。そして、実際の測定値との残差は  $|y_i - f(x_i)|$  となる。この残差は  $xy$  平面上での点  $(x_i, y_i)$  とグラフ  $y = f(x)$  の  $y$  軸方向の距離である。

次に、理論値から誤差の分散の推定値は残差の平方和

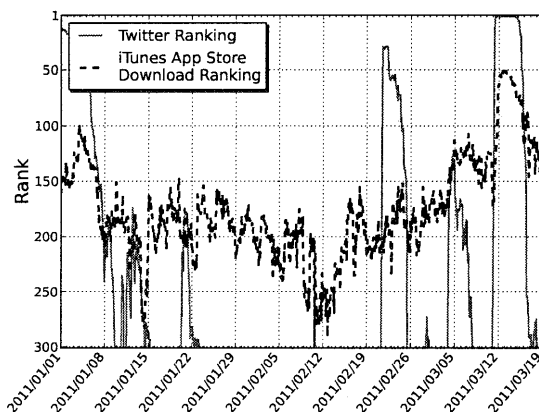


図 1 あるアプリケーションのスコア化によるつぶやき数、及び iTunes App Store のランキングの例 (iMovie, id: 377298193)

Fig. 1 Example of an application's ranking of both the number of tweet and the iTunes App Store using score (iMovie, id: 377298193)

$$J = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2)$$

で与えられる。よってこの  $J$  を最小とする想定分布  $f$  を定めれば良いということになる。

ここで想定される関数  $f(x)$  は既知の関数  $g(x)$  の線形結合であると仮定する。即ち

$$f(x) = \sum_{k=1}^m a_k g_k(x) \quad (3)$$

が成り立っていると仮定する。すると  $J$  を最小とする想定分布  $f$  を定めることは、 $a_k$  を定める事と同義となり、式 (2) は  $a_k$  を変数とする関数とみなすことができるので、 $J$  を  $a_k$  について偏微分したものをゼロと置き、得られた  $m$  個の連立方程式を解き、 $a_k$  を決定する。

#### 4.3.2 本研究における最小二乗法

本研究では特定期間のランキングの時間変化を検出するために Twitter のつぶやき数ランキング、iTunes App Store のランキング双方に対して最小二乗法を用いた。想定分布を下記に示す一次方程式 (4) とし、傾き  $a$  と切片  $b$  を求め直線近似を行った。

$$y = ax + b \quad (4)$$

ここで  $x$  は時間、 $y$  はある時間  $x$  のランキングである。 $a$ 、 $b$  は次式で求められる。

$$a = \frac{n \sum_{k=1}^n x_k y_k - \sum_{k=1}^n x_k \sum_{k=1}^n y_k}{\sum_{k=1}^n x_k^2 - \left( \sum_{k=1}^n x_k \right)^2} \quad (5)$$

$$b = \frac{\sum_{k=1}^n x_k^2 \sum_{k=1}^n y_k - \sum_{k=1}^n x_k y_k \sum_{k=1}^n x_k}{\sum_{k=1}^n x_k^2 - \left( \sum_{k=1}^n x_k \right)^2} \quad (6)$$

なお、本研究ではある時間  $x_i$  から 6 時間分の値で直線近似を行い、その近似式をある時間  $x_i$  の近似直線としている。ここで  $i$  は時間のインデックスである。

$$y_i = a_i x_i + b_i \quad (7)$$

本研究で用いた  $a_i$ ,  $b_i$  を求める式を次に示す。

$$a_i = \frac{6 \sum_{k=i}^{i+6} x_k y_k - \sum_{k=i}^{i+6} x_k \sum_{k=i}^{i+6} y_k}{\sum_{k=i}^{i+6} x_k^2 - \left( \sum_{k=i}^{i+6} x_k \right)^2} \quad (8)$$

$$b_i = \frac{\sum_{k=i}^{i+6} x_k^2 \sum_{k=i}^{i+6} y_k - \sum_{k=i}^{i+6} x_k y_k \sum_{k=i}^{i+6} x_k}{\sum_{k=i}^{i+6} x_k^2 - \left( \sum_{k=i}^{i+6} x_k \right)^2} \quad (9)$$

これにより Twitter つぶやき数ランキング、及び iTunes App Store ランキングのある時間  $x$  の傾きを計算し、その傾きが  $a \leq -10$  であったものを急上昇とし検出する。なお、急上昇時に傾きがマイナスとなるのは、 $y$  軸をランキングの値としてとっているからであり、その値の減る割合が大きいほど急上昇と見なせるからである。

#### 4.4 傾きによる場合分け

各ランキングにおける傾きの値の正負によって以下の 3 通りに場合分けを行った。

##### Case 1

$Twitter a_i < 0$  かつ  $iTunes a_i < 0$   
(双方共に増加)

##### Case 2

$Twitter a_i \geq 0$  かつ  $iTunes a_i < 0$   
(iTunes App Store ランキングのみ上昇)

##### Case 3

$Twitter a_i < 0$  かつ  $iTunes a_i \geq 0$   
(Twitter つぶやき数ランキングのみ上昇)

また、Case 1 に関しては更に次の 3 通りに場合分けを行った。

##### Case 1-1

$Twitter a_i \leq 10$  かつ  $iTunes a_i \leq 10$   
(双方ともに急上昇)

##### Case 1-2

$-10 < Twitter a_i < 0$  かつ  $iTunes a_i \leq 10$   
(iTunes App Store ランキングのみ急上昇)

##### Case 1-3

$Twitter a_i \leq -10$  かつ  $-10 < iTunes a_i < 0$   
(Twitter つぶやき数ランキングのみ急上昇)

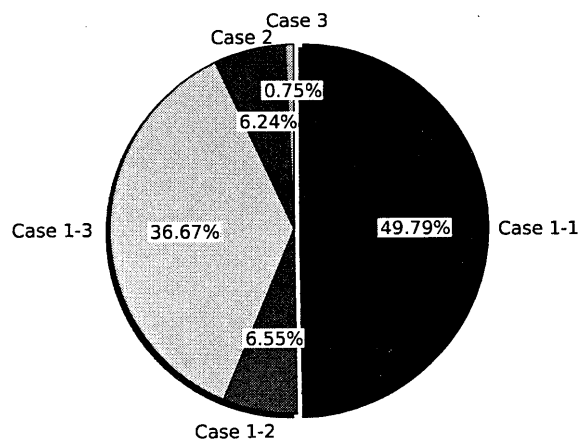


図 2 傾きによる場合分けの割合

Fig. 2 Case analysis by  $a_i$

#### 4.5 時間遅れ検出

あるランキングから急上昇が検出されたら、その検出された時間を基準とし、もう片方のランキングの  $\pm 24$  時間分の直線近似における傾きをみて、その最大値をとる時間までの差分をつぶやき数ランキングと iTunes App Store ランキングとの時間遅れとする。

$$Twitter d_i = \arg \max_{-24 \leq j \leq 24} (iTunes a_{i+j}) \quad (10)$$

$$iTunes d_i = - \arg \max_{-24 \leq j \leq 24} (Twitter a_{i+j}) \quad (11)$$

ここで、 $d_i$  はインデックス  $i$  におけるつぶやき数ランキング上昇と iTunes App Store ランキング上昇の時間遅れ、 $\arg \max$  は引数の最大値を取るようなインデックスの値を返す関数である。

この時間遅れ  $d_i$  は、Twitter のつぶやき数ランキングが急上昇する立ち上がりの時間を基準とし、それに対して iTunes App Store のランキング急上昇が何時間後に急上昇するかを表している。これら時間遅れについて度数分布図を作成し、その分布を観察した。

また、 $d_i = 0$  からの累積度数をとり、その累積度数分布図を作成した。この図より、つぶやき数ランキングが急上昇してから何時間後に iTunes App Store ランキングが立ち上がるのかが観測できる。

#### 5. 計算結果

統計の結果、つぶやき数ランキング、あるいは iTunes App Store ランキングのいずれか、もしくは双方が急上昇を見せたアプリケーション数は 1235 個であった。その中でつぶやき数ランキング、あるいは iTunes App Store ランキングのいずれか、もしくは双方が急上昇を見せた回数は 28080 回であった。

##### 5.1 傾きの分布

4.4 で述べた場合分けを行った結果とその割合を図 2 に示す。この図より、4.4 で述べた場合分けのうち Case 1-1、つまり

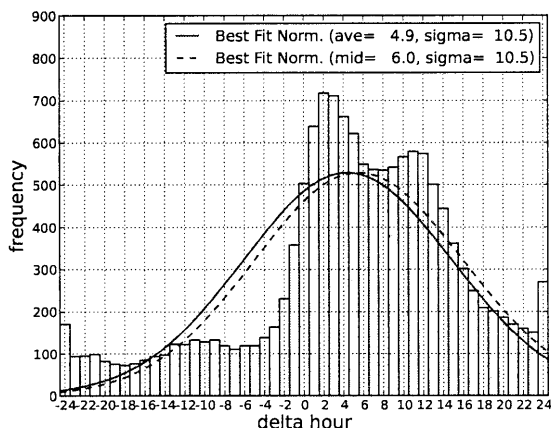


図3 時間遅れの度数分布図と近似正規分布曲線

Fig.3 Histogram of  $d_i$  and Best Fit Normal Distribution

つぶやき数のランキングの急上昇と iTunes App Store ランキングの急上昇は約 50%の確率で 24 時間以内に起こることが分かった。

## 5.2 時間遅れの分布

つぶやき数ランキングと iTunes App Store ランキングの間の時間遅れの度数分布図と、その平均値（実線）または中央値（破線）と標準偏差を使った正規分布曲線を図 3 に示す。時間遅れの平均値は+4.90 時間、中央値は+6 時間、最頻値は+2 時間、標準偏差は 10.52 時間であった。図 3 中の  $\pm 24$  の部分の度数が多いのは  $\pm 24$  時間以上ずれているものが全て  $\pm 24$  の部分に丸まってできた打ち切りによるものと思われる。この部分を例外値とし除外すると、平均値は+4.88 時間、中央値は+6 時間、最頻値は+2 時間、標準偏差は 9.82 となった。

また、 $d_i < 0$ 、 $d_i = 0$ 、 $d_i > 0$  の割合はそれぞれ、21.75%、3.60%、74.65%となった。これにより 5.1 で述べた Case1-1 のうち、約 3/4、即ち

$$49.79\% * 74.65\% = 37.17\% \quad (12)$$

が全体のうち、つぶやき数のランキングが先行して急上昇し、それに追従して iTunes App Store ランキングが急上昇する、という様子がわかった。

また、 $d_i = 0$  からの時間遅れの累積度数を、母数を  $0 \leq d_i \leq +24$ 、 $-24 \leq d_i \leq 0$  全体としたときの割合の詳細な数値を表 1、及び表 2 に示す。

表 1 より、Twitter のつぶやき数ランキングが急上昇してから+6 時間以内に 40.25%、+12 時間以内に 70.72%、+18 時間以内に 89.61%の割合で iTunes App Store ランキングが上昇してきている様子が見られた。

## 6. 考察と課題

5. の結果より、iTunes App Store のランキングと Twitter のつぶやき数ランキングは、約 50%ほどの割合で、24 時間以内に双方ともに上昇するという結果が得られた。また、Twitter のつぶやき数ランキングが急上昇してから iTunes App Store

表 1 累積度数の割合 ( $d_i \geq 0$ )Table 1 Percentage of Cumulative Frequency of  $d_i$  from 0 ( $d_i \geq 0$ )

$d_i$	Percentage	$d_i$	Percentage
$\pm 0$	4.60%	$\pm 0 \sim +13$	75.30%
$\pm 0 \sim +1$	10.44%	$\pm 0 \sim +14$	79.36%
$\pm 0 \sim +2$	16.99%	$\pm 0 \sim +15$	82.67%
$\pm 0 \sim +3$	23.49%	$\pm 0 \sim +16$	85.42%
$\pm 0 \sim +4$	29.55%	$\pm 0 \sim +17$	87.70%
$\pm 0 \sim +5$	35.23%	$\pm 0 \sim +18$	89.61%
$\pm 0 \sim +6$	40.25%	$\pm 0 \sim +19$	91.44%
$\pm 0 \sim +7$	45.16%	$\pm 0 \sim +20$	93.14%
$\pm 0 \sim +8$	50.05%	$\pm 0 \sim +21$	94.70%
$\pm 0 \sim +9$	55.01%	$\pm 0 \sim +22$	96.16%
$\pm 0 \sim +10$	60.18%	$\pm 0 \sim +23$	97.53%
$\pm 0 \sim +11$	65.47%	$\pm 0 \sim +24$	100.00%
$\pm 0 \sim +12$	70.72%		

表 2 累積度数の割合 ( $d_i \leq 0$ )Table 2 Percentage of Cumulative Frequency of  $d_i$  from 0 ( $d_i \leq 0$ )

$d_i$	Percentage	$d_i$	Percentage
$\pm 0$	14.19%	$\pm 0 \sim -13$	70.57%
$\pm 0 \sim -1$	24.29%	$\pm 0 \sim -14$	73.36%
$\pm 0 \sim -2$	30.81%	$\pm 0 \sim -15$	75.67%
$\pm 0 \sim -3$	35.44%	$\pm 0 \sim -16$	78.42%
$\pm 0 \sim -4$	39.33%	$\pm 0 \sim -17$	80.70%
$\pm 0 \sim -5$	42.69%	$\pm 0 \sim -18$	82.61%
$\pm 0 \sim -6$	46.05%	$\pm 0 \sim -19$	84.44%
$\pm 0 \sim -7$	49.18%	$\pm 0 \sim -20$	87.14%
$\pm 0 \sim -8$	52.54%	$\pm 0 \sim -21$	89.70%
$\pm 0 \sim -9$	56.29%	$\pm 0 \sim -22$	92.16%
$\pm 0 \sim -10$	59.90%	$\pm 0 \sim -23$	95.53%
$\pm 0 \sim -11$	63.66%	$\pm 0 \sim -24$	100.00%
$\pm 0 \sim -12$	67.10%		

のランキングが急上昇する割合は約 37%、その時間差は平均値で 4.9 時間後、中央値で 6 時間後となる傾向が見られた。

図 3 において、それぞれの平均値と標準偏差を使った正規分布曲線を図中の赤破線で示す。この曲線は、時間遅れの度数分布図に近い曲線と見ることができる。よって、つぶやき数のランキングと iTunes App Store のランキングの時間遅れはある程度において正規分布に近い分布になるのではないかと予測される。また、同図から、 $+2 \leq d_i \leq +3$  時間あたりにピークがあり、その後  $+10 \leq d_i \leq +12$  時間あたりにもうひとつのピークがある。この傾向より、注目度が高いアプリケーションに関して、いち早く iTunes App Store のランキング変動に反応するパターンと、半日程度の時間をかけてユーザに広まっていく、という 2 つの情報伝播のパターンがあるのではないかと考えられる。更には、 $+12 \leq d_i \leq +23$  の範囲で、 $d_i$  が増加するに従ってその度数は減っていることが確認できる。これより、Twitter において情報が広まりきるのはほぼ 24 時間以内に完了してしまうのではないかと予測される。

次に、表 1 と表 2 を比べると、 $0 \leq d_i \leq |6|$  程度の範囲にお

いて、やや表 2 ( $d_i \leq 0$ ) の方が表 1 ( $d_i \geq 0$ ) に比べて、上昇率が急になっている。このことから、iTunes App Store のランキングが Twitter で話題になるのは、その逆の場合に比べて非常に早いのではないかと、ということが推測できる。

次に、本研究について、また本研究を通じて発見した課題について述べる。まず、本研究は現時点で 3ヶ月弱程度、アプリケーション数にして 1235 種類のサンプル数しか取っていない。そこで、この期間を増やして 1 年間程度として調査するとどうなるかを確認したい。また、この期間には正月や東日本大震災など様々な出来事が含まれている。このような特異な出来事が起きた際、人は情報に敏感になると予測できる。その予測を持って、そのような出来事の前後の期間を精査すると、全体と比較してどうなるのか、という比較検証を行いたい。今回の研究では単純なつぶやきの数のみを見ている。ゆえにつぶやいたユーザがどのようなユーザであろうと、またその本文がどのような内容であろうとその情報は無視されている。しかし実際はフォロワー数が多い、有名人である、また悪評のつぶやきなど、そのつぶやきの重要性は様々であると考えられる。そこでその情報を考慮してつぶやきの重み付けをしたらどうなるか、ということも研究課題としてあげられる。

今回は急上昇の直線近似、またその傾きの閾値、近似の窓などいくつか定義を行った上で統計処理を行った。このあたりの数値も期間を長く取り、まずはその値に対して統計処理を行うことで、妥当な値を計算することも必要だと思われる。また直線近似には 6 時間の窓の最小二乗法を用いたが、最小二乗法による回帰直線の傾きの推定値は、より小さい方向にずれてしまうことが知られている [10]~[12]。本研究の場合だと、窓より小さな範囲内にピークがあった際に、その傾きを正しく検出できない可能性がある。ゆえに、直線近似の方法を再考慮する必要がある。

## 文 献

- [1] 総務省, “平成 23 年度版情報通信白書,” Aug. 2011. <http://www.soumu.go.jp/johotsusintokei/whitepaper/>
- [2] A. Java, X. Song, T. Finin, and B. Tseng, “Why we twitter: Understanding microblogging usage and communities,” Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007, pp.56–65, Aug. 2007.
- [3] J. Bollen, H. Mao, and X.-J. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol.2, pp.1–8, March 2011. <http://arxiv.org/pdf/1010.3003v1>
- [4] 白木敦夫, 矢野幹樹, 酒井佑太, 小澤俊介, 杉木健二, 松原茂樹, 河口信夫, “モバイルアプリケーション推薦のための twitter 発言者の状況の推定,” 情報処理学会マルチメディア、分散、協調とモバイル (DICOMO) シンポジウム論文集, pp.251–257, July 2010.
- [5] “Twitter - public timeline”. [https://twitter.com/public\\_timeline](https://twitter.com/public_timeline)
- [6] Apple, “itunes search api”. <http://www.apple.com/itunes/affiliates/resources/documentation/itunes-store-web-service-search-api.html>
- [7] ReadWriteStart, “Hacking the iphone app store’s ranking algorithm,” Feb. 2010. <http://www.readwriteweb.com/start/2010/02/iphone-appstore-ranking-algorithm.php>
- [8] faberNovel, “Ratp premium: Subway & bus in paris official”. <http://itunes.apple.com/us/app/ratp-premium-subway-bus-in/id313907721?mt=8>
- [9] “fabernovel.com”. <http://www.fabernovel.com/>

- [10] 小池文人, “最小二乗法を使わない簡単な直線回帰のやりかた”. <http://vege1.kan.ynu.ac.jp/nakaumi/turnover/regression.html>
- [11] R.R. Sokal and F.J. Rohlf, Biometry. 3rd edition., Freeman, 1995.
- [12] 粕谷英一, 生物学を学ぶ人のための統計のはなし 一君にも出せる有意差一, 文一総合出版, 1998.